Tampere University

Petri Salminen

TOWARDS SMART BUILDINGS

Utilizing sensor data and building information model in a multi-purpose environment

Master's Thesis
Faculty of Engineering & Natural Sciences
Hannele Väyrynen
Kalle Kähkönen
Timo Lehtonen
November 2020

ABSTRACT

Petri Salminen: Towards Smart Buildings – Utilizing sensor data and building information model in a multi-purpose environment
Master of Science Thesis
Tampere University
Master's Degree Programme in Information and Knowledge Management
November 2020

Smart buildings are built structures, which utilize information technological solutions to share information between different parts of the building and to automize different actions, such as air conditioning, heating and light conditions.

The objective of this master's thesis is to provide insight on how sensor data and the building information models could be utilized in a multi-purpose environment. Additionally, how sensor data and business information model can be combined practically. A case building is used in order to study multi-purpose environments. The case building is located in one of the Finland's largest cities and it contains a school, a maternity clinic and a library, for example. The case building has some characteristics of a smart building, such as sensors and automation, but they are not fully utilized. The thesis is performed as a part of 4Apis project, which is funded by Business Finland. The thesis consists of literature review, in which the literature surrounding the combination of sensor data and building information models is studied, and an empirical interview research, in which the staff and other stake holders of the case building are interviewed.

Many use-cases for the combination of sensor data and building information model was found from the literature. Literature use-cases were categorized into four different themes: people in 3D space, structural health monitoring, safety and sustainability.

From the literature, many ways of combining sensor data and building information model were found. For example exporting the BIM model into a more sophisticated 3D model. Although a majority of them were so-called ad hoc solutions, one model was found from the literature, which was designed for general usage, but there were no signs at all of its use in practice.

The case building had sensor data, which can be utilized with data analytics. The data was retrieved in a raw spreadsheet format, which was not utilizable as such, because of the structure of the data. By preprocessing, the data was processed by the study group into a form, from which it was possible to generate easily interpretable charts by the study group.

By analyzing the sensor data, many phenomena could be verified. The class rooms of the case building were empty in the evenings, at night and on the weekends. This means that a majority of time, the classrooms were empty. From the temperature sensors it was noticed, that the temperature in a third floor class room rose higher during the days, when compared to a class room which was located in the second flood. A more in depth analysis of the sensor data is necessary during further research.

In order to retrieve more qualitative data about the benefits of sensor data and building information models for the building users, three people, the principal of the case building, the energy supervisor of the city and the service coordinator of the case building, were interviewed. The interviews were conducted in a semi-structured fashion, with an emphasis on free discussion. The interviewees' thoughts about the combination of sensor data and building information model were discussed in the interviews. There were three themes that were recognized from the interviews: efficacy, safety and wellbeing. Especially the emergency operations and energy efficiency were represented in the interviews. In both, the interviews and the literature, highlighted safety, sustainability and efficacy. Because of the small sample size of the interview study, it is recommended that the subject should be studied with a larger sample size.

This thesis offers multiple views to the combination of sensor data and building information model, especially to the field of building technology In the future, the subject should be researched for example with a more in-depth data analysis or with empirically testing the solutions.

Keywords: smart building, sensor data, building information modeling
The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

TIIVISTELMÄ

Petri Salminen: Kohti älykkäitä rakennuksia — Sensoridatan ja rakennuksen tietomallin hyödyntäminen monitoimiympäristössä
Diplomityö
Tampereen yliopisto
Tietojohtamisen diplomi-insinööriohjelma
Marraskuu 2020

Älykkäät rakennukset ovat rakennuksia, jotka hyödyntävät tietoteknisiä ratkaisuja jakaakseen tietoa rakennuksen eri osien välillä ja automatisoidakseen eri toimintoja, kuten ilmastointia, lämmitystä ja valaistusta.

Tämän diplomityön tarkoituksena on selvittää, miten sensoridataa ja rakennusten tietomalleja voitaisiin hyödyntää monitoimitaloissa. Lisäksi selvitetään, miten sensoridatan ja rakennuksen tietomallin pystyy käytännössä yhdistämään. Työssä käsitellään monitoimitaloja case-rakennuksen avulla. Case-rakennus sijaitsee eräässä Suomen suuressa kaupungissa ja sisältää esimerkiksi koulun, neuvolan ja kirjaston. Case-rakennuksessa on älykkään rakennuksen piirteitä, kuten sensoreja ja automaatiota, mutta niitä ei hyödynnetä täysin. Työ on suoritettu osana Business Finlandin rahoittamaa 4Apis-hanketta. Työhön sisältyy kirjallisuuskatsaus, jossa käsitellään sensoridatan ja rakennuksen tietomallin yhdistämistä kirjallisuudessa, ja empiirinen haastattelututkimus, jossa haastateltiin case-rakennuksen henkilökuntaa sekä muita rakennuksen toiminnan tuntevia henkilöitä.

Kirjallisuudessa sensoridatan ja rakennuksen tietomallin yhdistämiselle nousi esille monia hyödyntämiskohteita. Nämä hyödyntämiskohteet pystyttiin jaottelemaan neljään eri teemaan: ihmiset kolmiulotteisessa tilassa, rakenteiden terveystarkkailu, turvallisuus ja kestävyys.

Sensoridatan ja rakennuksen tietomallin yhdistämiselle löytyi kirjallisuudesta monia eri tapoja. Esimerkiksi rakennuksen tietomallin lataaminen hienostuneempaan kolmiulotteiseen malliin. Vaikka suurin osa ratkaisuista oli ns. ad hoc -ratkaisuja, kirjallisuudesta löytyi myös yksi malli, joka oli suunniteltu yleiseen käyttöön mutta sen hyödyntämisestä käytännössä ei löytynyt viitteitä.

Case-rakennuksella oli käytössään sensoridataa, jota voidaan hyödyntää data-analytiikan avulla. Data saatiin raa'assa taulukkomuodossa, joka ei ollut sellaisenaan hyödynnettävissä analyysissä. Tutkimusryhmä sai datan esikäsittelyllä sellaiseen muotoon, josta heidän oli mahdollista luoda helposti tulkittavissa olevia kuvaajia.

Analysoimalla sensoridataa voitiin todentaa monia ilmiöitä. Case-rakennuksen luokkahuoneet olivat tyhjillään iltaisin, öisin ja viikonloppuisin, eli suurimman osan ajasta. Lämpötilasensorista huomattiin, että kolmannessa kerroksessa olevan luokkahuoneen lämpötila kasvoi huomattavasti enemmän, kuin toisessa kerroksessa olevan. Sensoridatan syvempi analyysi on tarpeen jatkotutkimuksessa.

Jotta saataisiin lisää ymmärrystä sensoridatan ja rakennuksen tietomallin yhdistämisen hyödyllisyydestä monitoimirakennuksen käyttäjille, haastateltiin kolmea henkilöä: case-rakennuksen rehtoria, kaupungin energiavalvojaa ja case-rakennuksen palvelukoordinaattoria. Haastattelut suoritettiin puolistrukturoituina, vapaata keskustelua korostavina haastatteluina. Haastatteluissa keskusteltiin, mitä haastateltaville tulee mieleen sensoridatan ja rakennuksen tietomallin yhdistämisessä. Haastatteluissa korostui kolme teemaa: tehokkuus, turvallisuus ja hyvinvointi. Erityisesti pelastustoimen ja energiatehokkuuden näkökulmat olivat haastatteluissa esillä. Sekä haastatteluissa, että kirjallisuudessa korostuivat turvallisuus, kestävyys ja tehokkuus. Haastateltavien pienen määrän vuoksi on suositeltavaa, että aihetta tutkitaan laajemmalla otoksella.

Tämä työ tarjoaa monipuolisia ja monialaisia näkökulmia sensoridatan ja rakennuksen tietomallin yhdistämiseen etenkin rakennustekniikan alalle. Jatkossa on kuitenkin syytä tutkia aihetta lisää esimerkiksi syvemmän data-analyysin tai ratkaisujen empiirisen testauksen muodossa.

Avainsanat: älykkäät rakennukset, sensoridata, rakennuksen tietomalli

PREFACE

This thesis was created as a part of 4Apis project, which is funded by Business Finland. I started working on the thesis in January 2020 and at first, I attended the 4Apis project meetings to get to know the people and the project. I found it very interesting and got to meet many interesting people. I would like to thank both the Business Finland, the whole 4Apis project and its participants. It was an honour to work with them and I am looking forward to co-operating in the future.

Of course, this thesis would not have succeeded without the support and feedback from the supervisors: Hannele Väyrynen, Timo Lehtonen and Kalle Kähkönen. Their expertise and wisdom guided me throughout the thesis. I would also like to thank the whole staff of Faculty of Management and Business, and especially Nina Helander and Jussi Myllärniemi. They recommended me for the master’s thesis worker role for this particular project.

Lastly, I would like to thank my fiancée Henna Ponkala. She has supported me through this whole project and helped me keep my mind together. I would not have succeeded without her. It cannot be stressed enough that this thesis would not have been possible without all the help I have received.

“Alone we can do so little, together we can do so much.”
-Helen Keller

Tampere, November 14th, 2020

Petri Salminen

CONTENTS

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

3D	Three dimensional
API	Application Programming Interface
BIM	Building Information Modelling
BIMSL	Building Information Modeling Sensor Language
BMS	Building Management System
DSL	Domain Specific Language
FINER	Feasible, Interesting, Novel, Ethical, Relevant
HVAC	Heating, Ventilation and Air Conditioning
IFC	Industry Foundation Classes
IoT	Internet of Things
SHM	Structure Health Monitoring
WiFi	Wireless Fidelity
WLAN	Wireless Local Area Network

1. INTRODUCTION

During the modern information era, it has become a trend to apply intelligence to everything from egg holders to highways. The build environment is no exception. Smart buildings along with smart materials and smart sensors have been researched for the last three decades (Buckman et al. 2014). Buildings can be enhanced with smart technology in order to achieve savings or increased efficiency. It is also said, that in built environment, digitalization is one of the most important changes since the industrial revolution (The Institution of Structural Engineers 2020).

The global warming and the climate change cannot be ignored and more and more restrictions about emissions are being made. This puts extra pressure on build structures, because buildings account for approximately 40% of global energy consumption and over 30% of the carbon dioxide emissions (Yang et al. 2014). It could be globally very beneficial, if the energy consuption and carbon dioxide emissions of the buildings would be lowered by just a little.

There has been studies in Finland about cities' energy efficiency programs (Sitra 2020). In Tampere, the sustainability of the built environment is the cornerstone of their energy efficiency programme and they have plans to utilize smart buildings in order to achieve carbon neutrality (City of Tampere 2013).

Smart buildings can provide comfort for those who visit the building on a regular basis. This comfort can lead to increased wellbeing as a whole, which in turn could provide for increased efficiency.

The aim of this thesis is to collect more knowledge about the combination of sensor data and building information models from the perspective of a multi-purpose environment. During the initial literature review, a research gap was found from the field of smart buildings; the approaches in literature are not aligned with the real-life applications. This thesis aims to narrow that research gap by introducing ways to utilize the combination of sensor data and building information models.

1.1 Smart buildings

Smart buildings have a close relation to building automation systems (Snoonian 2003). Building automation systems are systems, which control the Heating, Ventilation and Air Conditioning (HVAC) of a building (Kastner et al. 2005; Snoonian 2003). The market around building automation systems is revolved around the promise that they can both increase the user comfort and reduce the operation costs of the building (Kastner et al. 2005). According to Buckman et al. (2014), smart buildings are buildings which combine intelligence, enterprise, materials & design and control to achieve adabtiveness.

Already in 2003, Snoonian (2003) proposed smart buildings, which would have their Heating, Ventilation and Air Conditoining (HVAC) controlled automatically with an building automation system, which reads sensor data in real time and is connected to the internet with a standardized BACnet protocol. BACnet was originally introduced in 1995 already ("BACnet Overview" n.d.), but its use is still relevant in 2020 (Tang et al. 2020; Vieira et al. 2020).

Djenouri et al. (2019) conducted a meta-analysis focusing on applying machine learning in smart buildings. The found machine learning applications formed two distinct classes; occupant centric and energy/device centric. The applications in the former category includes occupancy analysis and activity recognition, while the latter includes energy profiling and optimizing and fault detection. (Djenouri et al. 2019)

1.2 Sensor data and building information model in schools

The literature was searched for studies which share multiple similarities with this study. The similarities that were searched for were: in the study, both the sensor data and the building information model are utilized together and separately and that the main focus is a public building. No studies were found, where the utilization of sensor data and building information model would have been studied or demonstrated in any way.

However, one study (Sangogboye et al. 2018) was found, which utilizes sensor data and building information model (vaguely) in a large education building. Sangogboye et al. (2018) did a occupancy analysis on a large educational building. The building has a floor area of 8500 m^2 and it was built in 2015. The case building included multiple sensors and solar panel system on the roof. They used Schneider Electric building management system (BMS) to control and monitor the building. The BMS also transfers the data into a time-series database. Sangogboye et al. (2018) monitored two rooms, a 125 m^2 study zone and a 139 m^2 classroom, and created a simulation in which the energy performance of the rooms was measured. Wifi and camera signal was used in order to produce the simulations. In the study, they managed to simulate the average occupancy with an error between 1% and 3%. The occupancy in the educational building was low at night.

1.3 The research objectives and methods

In order to respond to 4Apis project's knowledge need during the case of the case building, the focus of this thesis was chosen to be on combining sensor data with building information models.

Research question determines the uncertain or unknown subject, that the research wants to resolve or investigate. One way to evaluate a research question is to use FINER criteria (Cummings et al. 2013; Fandino 2019). FINER is an acronym of *Feasible, Interesting, Novel, Ethical* and *Relevant*. The criteria of the components of the FINER framework are represented in Table 1.

Table 1: FINER criteria to form a good research question. Adapted from (Cummings et al. 2013)

Component	Criteria
Feasible	The research question is answerable within the resources
Interesting	Answering the questions intrigues the researcher
Novel	Answering the question advances science
Ethical	There are no ethical show stoppers
Relevant	Relates to done research and simulates further research

The proposed the main research question is: *How can sensor data and building information model be utilized a multi-purpose environment?*

To support the main research question, supporting questions are researched too: *How can sensor data and building information models be combined in practice?*

The main question is answered more based on empirical data, whereas the supporting question is answered mostly based on literature. The evaluation of the research questions based on FINER framework is as follows:

Feasibility: Both research questions are answerable within the researcher's resources. The researcher's resources include literature access through the university, sensor data and interview data.

Interesting: The 4Apis project is interested in both of the research questions. Also, the researcher is interested in solving them. Multiple cities are interested in smart buildings, because built environment makes up the majority of the carbon emissions. One city, that is currently investigating smart buildings is Tampere. Tampere is interested in smart cities mainly because of climate awareness. (City of Tampere 2013)

Novel: During the preliminary literature research, no single answer to the questions could be found.

Ethical: Although there are multiple different aspects to research ethics (Israel and Hay 2020), not a lot of them concern the choosing of the research question. If the ethicality of the research question would evaluated with whether it does more good than harm, I would argue that the chosen research question would be considered ethical.

Relevant: The research questions are relevant for 4Apis project and the case building, because the representitives of the case city acknowledged that they could use the sensor data and building information models in order to achieve multiple kinds of benefits. For example, the case city could have monetial savings and they could be able to serve the citizens of the case city better. Also, the results could be used in one of the forementioned projects concerning Finnish sustainable smart cities.

1.4 Practical limits

Practical limits are limitations on the research, which can be justified by practical means. In this research, the scope is narrowed from building information modeling to building information models, because the building information modeling processes for the case building are not known and because very much of the literature focuses on three dimensional (three spatial dimensions, the ones we move in) aspect of the building information modelling. This introduces some difficulty with determining the terms of the study.

In order to limit the scope of the data analysis, different machine learning methods and algorithms are not tested thoroughly. This could be a possible research subject for future research.

1.5 Structure of this report

This report starts with a Introduction in chapter 1. Introduction comprises of the background of the subject (the field of study) and the broad descriptions of the research goals, practical limits and this structure of this report. In chapter 2, the terms and the research field is narrowed down and more focused view is taken on the building information modeling, sensor data in built environment and how sensor data and building information model can be combined according to literature and what can be benefited from it.

Chapter 3 is dedicated to methodology and for explaining the scientific framework around the research. Methodology and the scientific framework explain how and why this research is conducted. Especially the research is described in detail from the initial literature review to the interview conducting. After methodology and the research description, in chapter 4, gathered research data is introduced and analyzed.

Chapters 5 and 6 are results and conclusions, respectively. The results introduce the analyzed results, such as the themes that were gathered from the interviews. Conclusions summarizes the whole research, acknowledges some shortcomings in the research and lastly proposes subjects for future research.

2. DATA FOR BUILDING INFORMATION MODEL-LING

Data, information, knowledge and wisdom all have multiple definitions. (Rowley 2007) Data are facts or observations, which have no meaning or value. With interpretation, context and/or processing, data can be enhanced into information. Furthermore when information is combined with understanding and capability it is possible to gain knowledge and with lots of knowledge and altering views on multiple subjects, wisdom. (Rowley 2007)

The value from the data is created from the information use (Lim et al. 2018). In the scope of combining sensor data and building information model, the added value comes from the extra information dimension(s), which are added on the top of the dimensions in the building information model's dimensions. This can also increase information density.

2.1 Building information modeling and model

Building information modeling (BIM) is revolutionizing the building industry, because BIM makes it possible for multiple parties to collaborate on the same building plans whilst enhancing quality control and risk management. (Epstein 2012). However, there is confusion over what building information modeling is (Epstein 2012; The Institution of Structural Engineers 2020). BIM can be an abbreviation of Building Information Modeling (Epstein 2012) or Building Information Model. In this research, the focus will be mainly on building information model, but in this section, building information modeling is considered. Basically, in this study as a whole, building information model is considered the three dimensional (the three spatial dimensions, the ones we walk in) part of the building information modeling.

In ISO 19650 standard, a standard which is commonly used when defining building information modeling, building information modeling is defined as *"use of a shared digital representation of a built asset to facilitate design, construction and operation processes to form a reliable basis for decisions"*. (International Organization for Standardation 2018). Therefore, building information modeling is not, by definition, limited to the 3D modeling of buildings. 3D modeling is the process, where the building is modelled in 3D on a computer. This is used to assist the design and the building process of the building. 3D modeling is still a large part of building information modeling.

In addition to the three dimensions of the 3D modeling, there can be more dimensions to the building information modeling. The complexity of building information modellng can be demonstrated with nD-modeling, where D stands for dimension and n describes the amount of dimensions (Lee et al. 2005). The first three dimensions are the spatial dimensions commonly marked with x, y and z in the commonly used cartesian coordinate system (Lee et al. 2005). These are the three spatial dimensions that we live and move in. The fourth dimension is time (Lee et al. 2005). Time can be used to see different construction phases even before the construction has started (Kunz et al. 2002). This should reduce risk, since mistakes and conflicts can be found early on in the construction project (Lee et al. 2005).

United Kingdom’s The Institution of Structural Engineers (2020) suggests, that terms over four dimensions (4D) should not be used, because there are controversies in the dimensions above 5, and the fifth dimension, cost, is not really a dimension, because it is not independent on the previous four dimensions; especially cost changes over time. Sixth dimension is sometimes referred as faculty management and sometimes as sustainability (The Institution of Structural Engineers 2020), which can lead to confusion because it is not commonly agreed upon.

2.2 Sensor data technologies in built environment

Literature provides multiple ways of utilizing sensor data technologies in built environment and certain special themes can be detected. For example, sensor data can be used in built environment by monitoring people in the building, monitoring the structure heatlh, monitor a gas contamination inside the building and monitor energy consumption. Next, we will take a look at these themes of sensor data technologies in built environment. The technologies have been assigned into four distinct themes: People in 3D space, Structural health monitoring, Safety and Sustainability.

The first theme that was detected is *People in 3D space*. This point of view takes advantage of the fact that built environment often, happens in 3D space which is more or less utilized by people. It is possible to monitor people positioning within a building using data from multiple Wireless Local Area Network (WLAN) access points (Leppäkoski

2015) Also, occupancy information can be obtained from a combination of stereo-vision cameras and WiFi (Wireless Fidelity). (Sangogboye et al. 2018)

Structural health monitoring is the second detected theme, although according to Smith (2016), sensor data interpretation is often thought as a subfield of structural health monitoring (SHM), but sensor data interpretation is not always about SHM. Sensor data interpretation in SHM is more mature, than for example, in asset management and it has been studied much more (Smith 2016). Sensor data can be used in order to monitor cracks in steel bars in buildings (Smith 2016)

The third theme that was detected is *Safety*. Sensor data about particles in the building can be used in order to track contamination spread around the building. (Sharma et al. 2019) This could be especially useful, if building is used to handle and/or store dangerous gasses. Another example about safety is that firefighters have little knowledge about the true conditions inside a burning building. Sensor technology has become more inexpensive, which makes it plausible to implement large scale sensor systems inside buildings. Using real-time sensor data, very complex decisions can be made based on the data in a case of fire emergency. (Han et al. 2010)

The fourth theme that was detected is *Sustainability*. Energy consumption is being reduced in buildings in order to promote sustainability in the built environment (Howell et al. 2014). And in the recent years, new technologies have risen, which reduce the buildings' negative impact to environment. (Petrova et al. 2019). Also constraints on the energy efficiency of buildings are increasing. This leads to buildings gradually transforming into smart spaces, which have sensors and analytics. (Chitu et al. 2019)

2.3 Combining sensor data with building information model

In this chapter, a closer look on literature is taken on the subject of combining sensor data with building information model. This is important, because the integration of BIM and sensor data simplifies the design process between the sensor data and physical objects. (Alves et al. 2017) Simplifying the design process between the sensor data and physical objects could lead to better placement of the sensors, which in turn could increase the data quality gotten from the sensors.

2.3.1 Difficulties combining sensor data with building information model

There are many difficulties with combining sensor data and building information model, like the difficulty of implementation and the lack of standardization. According to Alves et al. (2017), the core reason for the difficulty of integrating business information model and sensor data is that the right queries with which the data is read are difficult to determine. Also, the built environment is complex, Alves et al. (2017) adds.

According to Alves et al. (2017), most current integration approaches in the literature remain theoretical or are highly dependent on a specific domain. As a response to this problem, Alves et al. (2017) developed a domain specific language (DSL) called Building Information Modeling Sensor Language (BIMSL). During our research, no real-life use-cases was found from literature for the BIMSL (Alves et al. 2017). This can indicate that literature and real-life applications do not have shared approaches.

Industry Foundation Classes (IFC) allows adding sensors to the building information model. For example, temperature and humidity sensors are supported by IFC. However, the existing BIM software, in 2013, was not able to process them correctly. Existing BIM software is not designed to support sensor data. (Rio et al. 2013)

2.3.2 Benefits & applications gained from combining sensor data with building information model

In the construction phase, the combination of building information modelling and sensor data can be used to protect the workers from dust in real time. (Smaoui et al. 2018) According to Smaoui et al. (2018) this can be achieved by many means, such as worker location tracking, real-time dust monitoring, using the building information model for visualization. (Smaoui et al. 2018)

One application in literature is, that a timeline is presented for the users. The timeline represents different states of the building at different points in time. The users can change the view back and forth in the timeline and see the data change immediately. Also, the user can playback the timeline with 900 times speedup. (Aleksandrov et al. 2019) This kind of timeshift ability could make it easier for the user to notice large scale patterns.

Sensors for carbon dioxide in the rooms of a building can be used to automatically adjust the air conditioning systems. For example, if carbon dioxide levels would rise above 900ppm, the air conditioning would turn on and provide more fresh air to the room. (Bang et al. 2019) Aleksandrov et al. (2019) managed to provide interesting graphical visualization to the carbon dioxide concentrations across different rooms.

Also other types of sensor to use were found from the literature. For example, Chiang et al. (2015) argue that sensor data could be extracted from the power outlets in order to monitor energy usage patterns. Another example is Riaz et al.'s study (2017), in which BIM is used for visualization. They used an external CoSMoS -program for reading the sensor data and connecting it to the corresponding space. Their building information model was in a proprietary Revit format and the solution relied strongly on proprietary tools, such as Revit and Visual Studio). Proprietary means, that it is not freely distributable and often the user needs a license in order to use it.

2.3.3 Combining sensor data with building information model

In literature, the combination of building information modelling and sensor data can be achieved by many means. There is no one single way, that is proven to be the right way to do it. Alves et al. (2017) approach the combination of building information modelling and sensor data by building a domain specific language on top of them, whilst storing the sensor data on a remote data system. In their (2017) method, the sensor data is transmitted from the remote data storage as data streams. Alternative way to integrate sensor data into BIM data is to export the BIM into a 3rd party 3D engine, such as Unity 3d (mostly known for being used in video game development) and then fetch the sensor data from database to client machine, which runs the Unity 3d model. (Chiang et al. 2015) Both of these perspectives attempt to circumvent the problem with non-compatible building information modeling -software by moving either the sensor data (Alves et al. 2017) or both the building information model and sensor data (Chiang et al. 2015) to a alternate location. Also, Riaz et al. (2017) uses a external storage for the data. Riaz et al. (2017) opted in for a NoSQL database called MongoDB, which according to Riaz et al. (2017) offers better performance in terms of fastness. Riaz et al. (2017) also considered timeseries databases, such as InfluxDB and Informix, but rejected them because their performance was very poor in their tests.

3. THE SCIENTIFIC FRAMEWORK AND METHODOLOGY

In this chapter, the methodology and the scientific framework for the research is considered. This includes information such as how the research is conducted and the reasoning behind the methodological choices that were done during the research.

3.1 The scientific framework

The scientific framework is introduced through Saunders et al.'s (2019) research onion model. The research onion model consists of several layers, which provide a systematic way to approach the research in a whole. (Saunders et al. 2019)

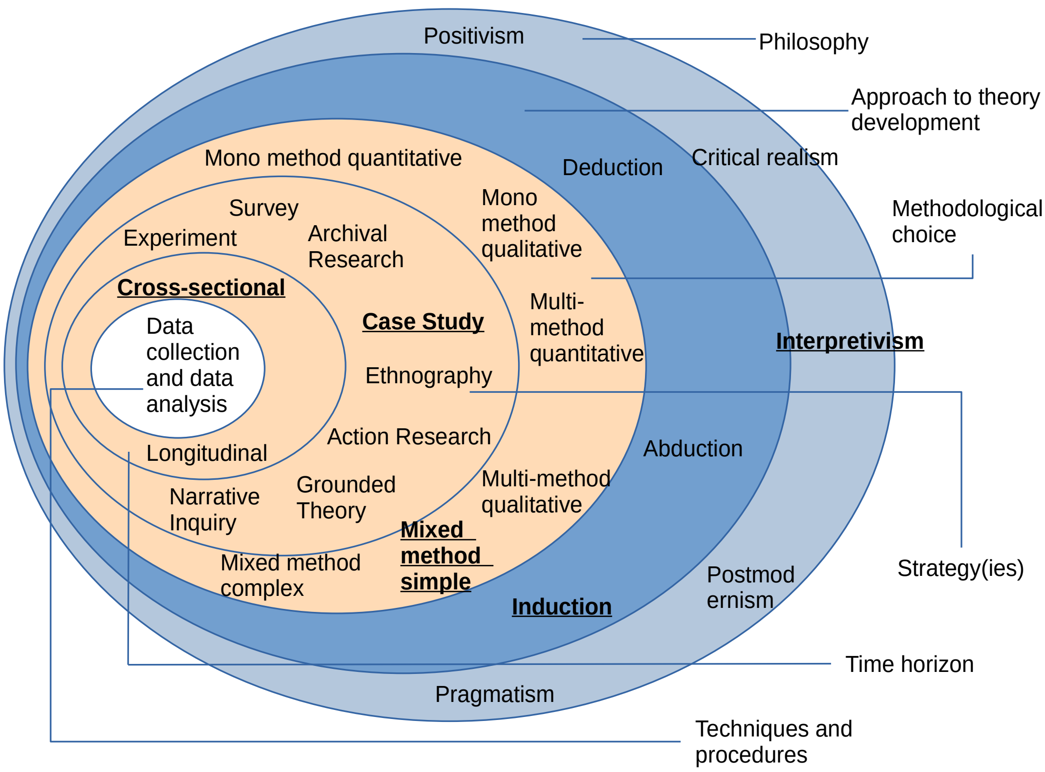

Figure 1: The research framework Adapted from Saunders et al. (2019)

In Figure 1, a research framework (Saunders et al. 2019), commonly called the onion model, is represented. The highlighted terms are the ones that are chosen for this study. The whole chapter 3 can be reflected to this model in order to have a complete picture of the scientific framework of the research.

3.1.1 Research philosophy

Research philosophy is the research' view, the beliefs and assumptions, on how knowledge is developed. (Saunders et al. 2019) This is essential, because according to Saunders et al. (2019), research is actually developing of knowledge in a particular field.

The research philosophy of the research is interpretivistic. The focus of interpretivistic research is to find new insights, understandings and interpretations on social worlds, contexts and phenomena (Saunders et al. 2019). People are active participants in the research, and their ideas in the interviews will be interpreted by the researcher.

3.1.2 Approach to theory development

The theory can be approached in three ways: deductive, inductive and abductive. Deductive approach is a way to approach the theory ”logically” as in *if premise A is true and it being true leads to conclusion B being true, then the conclusion B true*. Inductive, on the other hand, has a gap in the logic, which leaves room for interpretation. In inductive reasoning, if the premises are true, it indicates that the conclusion may be true. The third, abductive reasoning, often starts with a surprising fact, that is then explained through additional data or premises with testing. (Saunders et al. 2019)

Given the interpretivistic nature of the study, it would not be helpful to approach the theory from a deductive point of view. The gap for interpreting and reasoning is needed in the complex system. Also, the generalizing goes from specific to general is more common in inductive approach (Ketokivi and Mantere 2010; Saunders et al. 2019). Inductive approach allows the study to approach complex problems and provide insights about them.

3.1.3 Research design

Research design describes the data which is used in the research. The data can be, for example, raw data or interviews. Research design can be either quantitative, qualitative or mixed based on what kind of data is used in order to reach conclusions. (Saunders et al. 2019) The study has both quantitative sensor data and qualitative interview data, which are utilized in a mixed fashion. The data used in the research is discussed later in chapter 4.1.

3.1.4 Research strategy

Research strategy refers to the way that the researchers plans to reach their research goal. One of the research strategies is case study, which can be described as studying something (”case”) in a real-life setting. (Saunders et al. 2019) It is self-evident, that a case study is used, because the whole research resolves around the case building. The case study and the case building is discussed later in chapter 3.2.2.

3.2 Research description and methodology

During the early design phase of this thesis, the 4Apis project had a meeting, during which brainstorming was used as a tool to create ideas around the case subject. Brainstorming is a idea creation method, during which a high amount of ideas is brought up without inhibitation. (Law 2016) After the session, the ideas can be evaluated further.

The thesis is separated into two parts; the literature review and the case study. The literature review's main function in the thesis is to provide knowledge about the current state of the literature. The knowledge about the state of the literature is later used to reflect the findings from the case study.

There are two types of data gathered for the case study: qualitative data from semi-structural interviews and quantitative data from the sensors in the case building.

3.2.1 Literature review

During the initial literature review, more knowledge on the subject of combination of sensor data and building information model is gathered. This step is essential, because without understand of the current situation of the literature, it is very hard, if not impossible to draw reliable conclusions about the case study or the research data.

Throughout the literature review, Andor service is used to search for the articles. Andor provides a comprehensive access to most of the databases, which Tampere University has access to. The whole list of the databases used is found from the address https://libguides.tuni.fi/az.php. Andor has access to 415 databases, which is both a blessing and a curse; for even complicated search queries, results are found, but it is challenging to limit the search terms in a way, which provides a reasonable amount of results.

Happily, the search term *“sensor data” AND (BIM OR “building information model”)* with filters *“Full Text Online”* and *“Scholarly & Peer-reviewed”* yielded only 187 results. 187 is a very reasonable amount of articles for a starting point of a literature review. Out of these 187 articles, 37 was chosen based on the title. The titles which included both sensor data and building information model were included. Out of these 37 articles, 22 articles were chosen based on the abstract for the final literature review. The review was conducted by reading the remaining 22 articles, and taking notes about how to combine sensor data and building information model.

3.2.2 Case study

As the name suggests, case study focuses on a specific case. The case can be for example a person, a process or something completely different. Case study as a research method is way of understanding how the case interacts dynamically with its context in real-life setting. (Saunders et al. 2019) In this thesis, the case is the case building. The case study is conducted by cumulating understanding about case’s processes, people and values.

The research is conducted as a part of 4Apis project. 4Apis is owned by a consortion formed by University of Tampere and University of Helsinki together with multiple Finnish companies. The project is funded by Business Finland. The aim of the project is to enhance the organizations’ API (Application Programming Interface) abilities. Also, a concretic demo with a real customer-like demo subject will be built as a part of the project. The project started in the fall 2019 and will end in the spring 2021.

At first, the demo subject of the project was not defined, but the focus of the project started to lean towards build environments as the demo subject. ”How could APIs be used in buildings?” was the main question at the time. In January 2020 the case city, and more specifically the multifunctional school building in the city, was found to be an interesting demo subject. At the same time the topic was steared towards increasing the utilization of the spaces in the building. ”The Airbnb of office spaces” was also one idea. This idea behind “The Airbnb of office spaces” was a question whether it could be possible to rent office spaces of a building during the time it is not used for its primary use.

The case building locates in one of Finland’s largest cities. After the building of the case building started in February of 2017, the case building and kindergarten was opened gradually from the end of 2018 to the start of 2019. The case building has facilities for basic education, early childhood education, a library, youth services, school health care and maternity clinic. The common facilities between childcare, school health care and school enable closer cooperation between different professional groups in order to support the families. The objective would be, that a family would have the same nurse and doctor from the pregnancy through school.

The case building has a building information model available and additionally there is a pilot of running sensors in three rooms currently. Values, such as temperature, amount of people, amount of exhaust air and amount of carbon dioxide, can be read from the sensors. While combining sensor data with building information modelling does not concern APIs in any way, the 4Apis project showed interest on the subject. Also the representitive of the case building was interested in the idea of combining the sensor data with the building information model of the building.

3.2.3 Data analysis

The data analysis is conducted in order to find interesting facts from the sensor data of the case building. The case building is introduced in chapter 3.2.2.

The analysis is conducted in a explorative fashion, which means that the emphasis of the analysis is not to link two predetermined variables together or aggregate the data. Instead, the objective of the analysis is to get to know the data through constantly exploring the different variables and their connections to each others (multivariant analysis). To avoid common pitfalls of exploratory analysis, the results should always be approach with skepticism. (Hartwig 1979)

Apart from explorative analysis, the results of the analysises are tested with t-test. The p value is the product of t-test and it is used in order to measure the confidence about the result (Browne 2010). If the p value is smaller than the limit value (usually, 0.05 is considered a good limit value), then the result is not likely to have happened just because of random chance.

3.2.4 Interviews

The interviews are conducted as semi-structured interviews and they are sixty minutes in length each. Interview is a research method, in which a interviewer discusses a certain subject with the interviewee. The data resulting from a interview is what the interviewee said during the interview. First, ten minutes is reserved for establishing the goal and background of the interview for the interviewee, going through personal data processing practices and getting acquainted to each other.

Next, the basic knowledge about the interviewee is gathered via conversation. Fifteen minutes is reserved for this part. The goal is to extract information about the interviewee's organization and the interviewee's everyday life at work. This includes the interviewee's role and function in the organization, the interviewee's typical day and everyday processes and the interviewee's view about the interviewee's and the interviewee's organization's successes and failures.

The second part continues with a conversation about how much the interviewee currently works with data and data models, what kind of data models and what kind of experiences the interviewee has about working on data and data models. Also, the interviewee is asked about whether they have been working with the case building, why and what kind of experiences the interviewee has on working with the case building.

The focus on the third part of the interview is to converse about the 4Apis project and about the possibilities of the 4Apis project. The first question in this section is whether they know the 4Apis project beforehand, what kind of thoughts does the interviewee have about it and do they have any thoughts about building-bound data both in general and specifically with the case building.

The third part continues by the interviewer bringing up some hypotheses, visualizations and interesting observations. For each idea, the interviewee is asked for their opinion about the idea. The discussion may reveal the following: to whom is the idea important and why, what is needed in order for the idea to realize, to whom is the idea negative and why, how to anticipate, minimize or eliminate the risk, what should be done next and why.

If these subjects haven't come up yet in the previous discussion, they will be discussed here: indoor air solution (does the interviewee have information about indoor air optimization in other locations), the thematics of savings (are the expected savings economical, medical or ecological) and how could the combination of building information modeling and sensor data contribute to the Carbon neutral City project.

The interviews are ended with a revision of discussed matters. Also, the interviewee is asked about what idea sounds like the most promising one and what the next steps should be in their opinion.

4. RESEARCH DATA AND ANALYSIS

4.1 Sensor and interview data

There are two types of research data cosllected for this research: sensor data from the case building and interview data from different stakeholders of the case building. The stakeholders included the principal of the case building, the service coordinator of the case building and the energy supervisor of the city. More information about the interviews is covered in chapter 4.4.2.

The case building's sensor data was acquired from the representitives of the city. It was provided to us in a spreadsheets format. The spreadsheet file was automatically generated by the building automation system. It included multiple sensors that measured different conditions in three rooms of the case building. All in all, there were close to 590 000 data points in the data. One data point corresponded to one measurement of one sensor. The sensors, subsensors and their units are listed in Table 2.

Table 2: Sensors that are available in the case building

Sensor	Subsensors	Unit
People count	Total	pcs
	Count in	pcs
	Count out	pcs
	Battery voltage	V
	Signal strength (send)	dBm
	Signal strength (receive)	dBm
Exhaust air	Humidity	%
	Temperature	°C
	Battery voltage	V
	Signal strength (send)	dBm
	Signal strength (receive)	dBm
Exhaust air flow rate	Pressure difference	Pa
	Flow rate	l
	Signal strength (receive)	dBm
Indoor air quality	Carbon dioxide	ppm
	Humidty	%
	Pressure difference	Pa
	Temperature	°C
	Mass concentration (PM1.0)	µg/m^3
	Mass concentration (PM10.0)	µg/m^3
	Mass concentration (PM2.5)	µg/m^3
	Mass concentration (PM4.0)	µg/m^3
	Number concentration(PM0.5)	pcs/cm^3
	Number concentration(PM1.0)	pcs/cm^3
	Number concentration(PM10.0)	pcs/cm^3
	Number concentration(PM2.5)	pcs/cm^3
	Number concentration(PM4.0)	pcs/cm^3
	Pressure	hPa
	TVOC (Total Volatile Organic Compound)	ppb
	Typical particle size	µm
Incoming air	Humidity	%
	Temperature	°C
	Battery voltage	V
	Signal strength (send)	dBm
	Signal strength (receive)	dBm
Incoming air flow rate	Pressure difference	Pa
	Flow rate	l
	Signal strength (receive)	dBm

4.2 Tools used in the analysis

The analysis is run in a Jupyter environment. According to Yaniv et al. (2018) Jupyter notebooks provide a way to create well-documented and reproducible workflows. Jupyter notebook lets the user input code in multiple programming languages and the notebooks are meant for exploration and knowledge sharing. This is especially useful, because the analysis and the results are shared with the 4Apis project.

Together with the Jupyter notebook, Python programming language is used. The main reason for the choosing of Python is that the researcher has prior knowledge about its usage for data analysis. Especially the Pandas library is well-known by the researcher, which makes it faster to conduct explorative analysis.

The tools used are beneficial for the research, because they both 1) made the research easier to conduct and 2) made the research easier to replicate. Python is a very common programming language, which is used in multiple universities, such as Tampere University and the University of Helsinki.

4.3 Preprocessing the sensor data

In the preprocessing, a separate Jupyter Notebook was created called Preprocess.ipynb. Its purpose is to take the raw spreadsheet file of the sensor data and apply multiple operations to the dataset in order to make it easier to process at later time. The preprocessing notebook utilized Python Pandas library greatly.

```
import pandas as pd
FILENAME = "../data/data.xlsx"
df = pd.read_excel(FILENAME, header=3)

# Drop unneeded column
df.drop('Unnamed: 2', axis=1, inplace=True)

# Drop rows that contain missing values
df.dropna(inplace=True)

# Convert time from “13.1.2020” to “13.01.2020”
df['Aika'] = df['Aika'].apply(zeropad)

# Convert time into a datetime object
df['Aika'] = pd.to_datetime(df['Aika'], format="%d.%m.%Y %H:%M:%S")
```

Code 1: Basic preprocessing for the sensor data from the case building

In Code 1, the initial preprocessing steps are shown. First, the file is read and stored into a Pandas dataframe. The spreadsheet file had some kind of unnecessary information on the first two rows, so they can be removed in the reading phase already. There was also an unneeded column, which could be removed right away. Missing values in the dataset can make the analysis more difficult, so the next step is to remove rows, with missing values. Also, the column ‘Aika’ (time in English) proved to be difficult, because it was stored as a normal string without zeropadding in dates and months. The zeropad function is included in Appendix A: Zeropad Function.

When working with Pandas dataframes, the `Dataframe.head()` function can be used to print a table of the contents of the first rows of the dataframe. In Table 3, the first rows can be seen. By looking at the first rows, it is easy to comprehend the overall shape of the data. Currently, one row corresponds to one sensor measurement. In order to create analysis, of which time is the x-axis, we will have to “flip” the data so that the index is a datetime object and columns have the values for the corresponding sensor.

Table 3: The first 5 rows of the case building's sensor data

	Sensor	Subsensor	Unit	Time	Value
0	Exhaust air	Humidity	%	2020-01-28 13:30:00	41.2
1	Exhaust air	Temperature	°C	2020-01-28 13:30:00	20.6
2	Exhaust air flow rate	Pressure difference	Pa	2020-01-28 13:30:00	-4.2
3	Exhaust air flow rate	Flow rate	slm	2020-01-28 13:30:00	119.7
4	Incoming air flow rate	Pressure difference	Pa	2020-01-28 13:30:00	3.4

The time series form of the same data can be achieved by `Dataframe.groupby()` function. In Code 2, the operations needed to flip the dataset into a timeseries dataset are represented. First, the 'Sensor' and 'Subsensor' are merged into one. This is because grouping by two columns is much more complicated than grouping by one. Second, the groupby is ran. Groupby is a operation, which aggregates the data based on some grouping logic. In Code 2, the data is grouped by 'Sensor', and then means for every 10 minute period is calculated.

```
# To avoid duplicate sensor names,
# lets merge 'Sensor' and 'Subsensor' into 'Sensor'
df_merged = df
df_merged['Sensor'] = df['Sensor'] + '_' + df['Subsensor']

df_t = df_merged.set_index('Time')
df_t_weird = df_t.groupby('Sensor').resample('10min').mean()

df2 = df_t_weird.reset_index().pivot(index='Time', columns='Sensor',
values='Value')
```

Code 2: Operations ran in order to achieve timeseries dataset

At this point, the dataset looks absurd as is presented in Table 4. Happily, as shown in Code 2, the dataset can be pivoted, so that time is the index and the values are in columns.

Table 4: Output of df_t_weird.head(). The data looks weird right now.

		Value
Sensor	**Time**	
PeopleCount_Total	**2020-01-01 11:30:00**	0.0
	2020-01-01 11:40:00	0.0
	2020-01-01 11:50:00	0.0
	2020-01-01 12:00:00	NaN
	2020-01-01 12:10:00	0.0

After this preprocessing, the dataset looks as shown in Table 5. Note, that the sensors other than “IncomingAir_Humidity” and “Tuloilma_IncomingAir_Temperature” are omitted, because the 38 columns would have been hard to fit into a single page.

Table 5: Example of the preprocessed data

Sensor	IncomingAir_Humidity	IncomingAir_Temperature
Time		
2020-01-01 11:30:00	30.05	19.2
2020-01-01 11:40:00	29.95	19.2
2020-01-01 11:50:00	29.80	19.3
2020-01-01 12:00:00	29.60	19.3
2020-01-01 12:10:00	29.45	19.3

Now, it is very easy to chart any sensors on a time-series to see how to sensor value changes over time. This, however comes with a slight loss of accuracy in the sensor value. In this study, this lost accuracy is not of concern.

4.4 Results from the analysis

4.4.1 Results from the data analysis

In the case building, there are three rooms with sensors. The locations of the rooms are retrievable from the IFC model of the building. The locations relative to the IFC model's origo are listed in Table 6. The rooms C3032 and A3060 are located in third floor, while the room A2105 is in the second floor.

Table 6: The relative coordinates of the rooms wi th sensors

Room	X (m)	Y (m)	Z (m)
A2105	70.15	109.55	45.50
C3032	81.96	45.61	47.50
A3060	81.75	88.08	47.50

During a 4Apis project meeting, the representatives of the city made a hypothesis, that the rooms are not occupied in the evenings and on the weekends. In Table 7 and Figure 2 it can be seen that the hypothesis about the weekend occupation is probably true; the average number of persons in the room on both saturday and sunday is 0,00 compared to the 6-12 of working days. To test this, a T-test was calculated in order to test a hypothesis, that the two sets; weekend and not weekend have the same mean occupancy. The code for this comparison is shown in Code 3. The calculated p-value was $1.2975834373889216 \cdot 10^{-74}$. This is extremely small compared to a commonly used $\alpha=0.05$. Therefore, we can conclude that the weekend and working days do not have the same means.

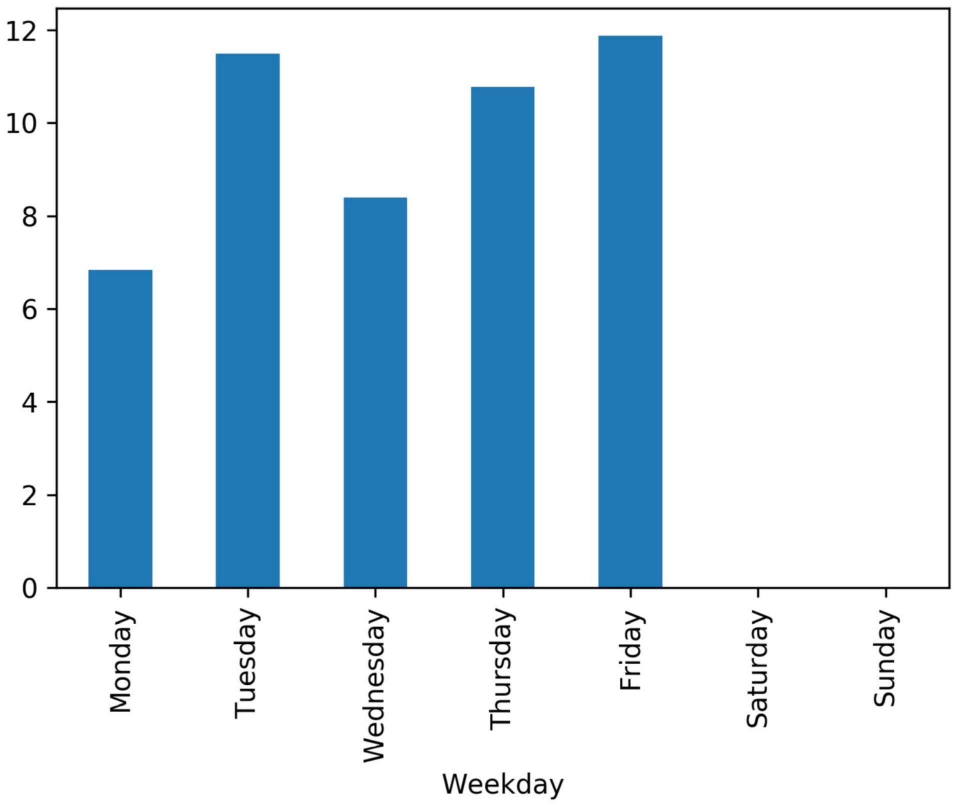

Figure 2: Average number of persons in the room by weekday

Table 7: Average number of persons by weekday

Weekday	Average number of persons
Monday	6,84
Tuesday	11,49
Wednesday	8,4
Thursday	10,78
Friday	11,88
Saturday	0,00
Sunday	0,00

```
import pandas as pd
from scipy.stats import ttest_ind

df = pd.read_csv('../data/Preprocessed.csv')
df['Aika'] = pd.to_datetime(df['Aika'])

is_weekend = df['Aika'].dt.weekday > 4
week = df[~is_weekend]['Henkilömäärä_Kokonaismäärä']
weekend = df[is_weekend]['Henkilömäärä_Kokonaismäärä']
ttest = ttest_ind(week, weekend, equal_var=False)
print(f"p-value, when comparing week and weekend: {ttest[1]}")
```

Code 3: Program, which reads the preprocessed file and calculates the T-tests, when comparing week and weekend datasets

Table 8: Average number of people in a given hour of a day

Starting hour	Average number of people
0	0.00
1	0.00
2	0.00
3	0.00
4	0.00
5	0.00
6	0.02
7	1.64
8	9.04
9	11.56
10	15.08
11	14.32
12	15.86
13	19.03
14	17.07
15	13.80
16	13.37
17	12.93
18	10.55
19	8.07
20	6.12
21	3.45
22	1.72
23	0.00

However, the hypothesis about evening usage (that is, that there is no evening usage) is not so clearly predictable by looking at Table 8, which has a bar graph of the average number of people in the room at given hour. This is shown in and Figure 3. However, when Code 4 was run, it showed that the occupancy level between 8 and 17 o'clock and occupancy level between 17 and 22 o'clock are certainly not the same (p-value= $5.83348598867209 * 10^{-20}$).

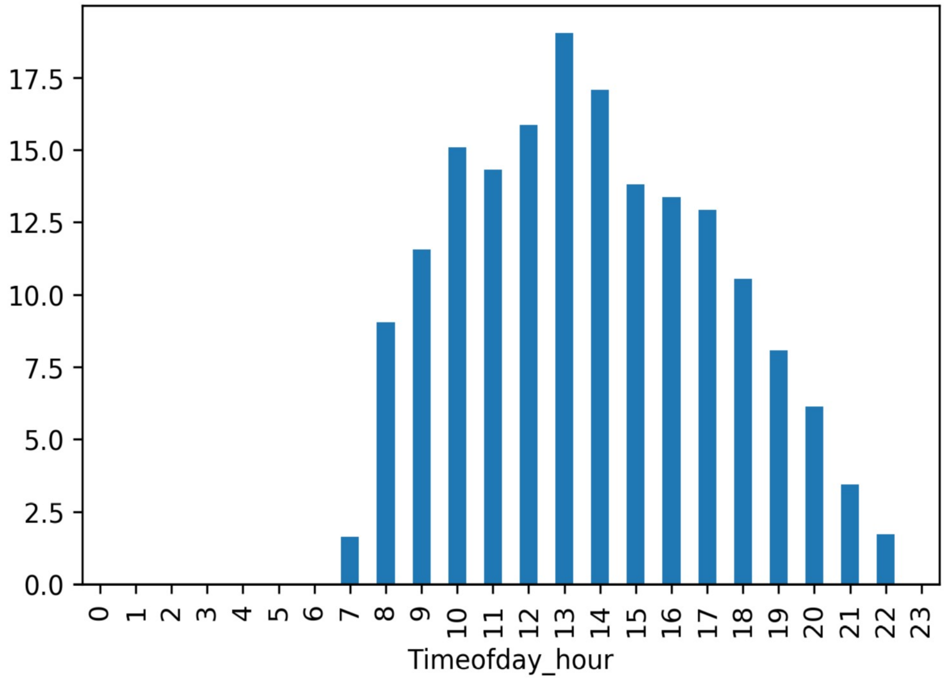

Figure 3: Average number of people in a given hour of a day

```
import pandas as pd
from scipy.stats import ttest_ind

df = pd.read_csv('../data/Preprocessed.csv')
df['Aika'] = pd.to_datetime(df['Aika'])

day_lower = df['Aika'].dt.hour > 7
day_upper = df['Aika'].dt.hour < 17
# Day is between 8 and 17
is_day = day_lower & day_upper

evening_lower = df['Aika'].dt.hour > 16
evening_upper = df['Aika'].dt.hour < 22
# Evening is between 17 and 22
is_evening = evening_lower & evening_upper

daytime = df[is_day]['Henkilömäärä_Kokonaismäärä']
evening = df[is_evening]['Henkilömäärä_Kokonaismäärä']
ttest = ttest_ind(daytime, evening, equal_var=False)
print(f"p-value, when comparing day and night:    {ttest[1]}")
```

Code 4: Program, which reads the preprocessed file and calculates the T-tests, when comparing daytime and evening

A hypothesis was made, that it is warmer in the upstairs rooms, because heat rises up. In Figure 4, it is shown that the curve with the class on the second floor is significantly lower than the third floor's rooms'.

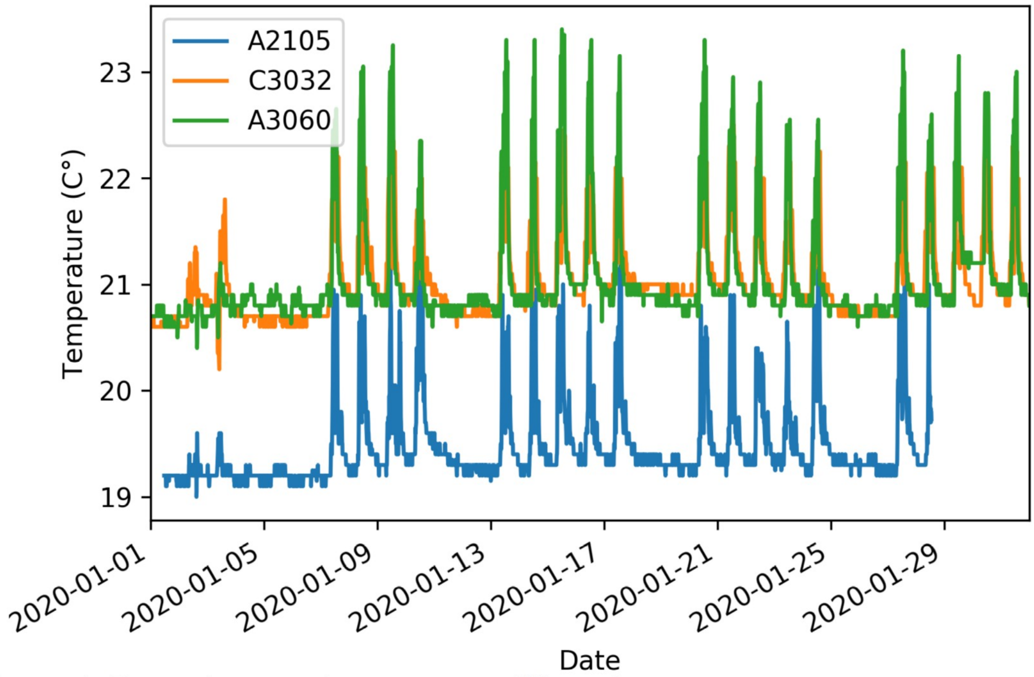

Figure 4: Room temperatures across different rooms

4.4.2 Results from the interviews

Three expert interviews were held to acquire rich qualitative data about the case building's case and about the combining of sensor data and building information model. The experts were chosen with a free discussion between 4Apis project members. The factors that were considered in the process were whether the interviewee is available and willing to participate and what is the estimated probability of getting interesting viewpoints from the interview.

The chosen interviewees were the principal of the case school, energy supervisor from the city and the service coordinator of the case building.

4.4.3 The first interview

When asked about the benefits and the possible benefits of combining sensor data and building information model, the principal of the case school mentioned staff wellbeing as one of the main benefits.

According to the principal of the case school, there is a phenomena called “mold anxiety”, which is general anxiety towards the indoor air quality of public facilities and especially schools in Finland. Sensor data could be used to produce facts, which can be used to reduce the said anxiety. According to the principal of the case school, this same logic applies to tax payers’ thoughts about the school; with fact-based information, it is easy to argue why some improvements are done and some are left unchanged.

The principal of the case school had some educational ideas as well. She stated that the sensor data could be used as a tool for constructivistic learning. For example, the data from the temperature and humidity sensors could be used in a physics class. According to the principal of the case school, bringing the study subject closer to the students could make hard subjects easier to grasp, because the subjects would be tied to the students’ every day environment. Also, according to principal of the case school, it could possibly improve the results with students with a learning difficulty. It could be beneficial, if the students could get involved with the process of improving the school with a tool like simple sensor analysis, the principal added.

When asked about how the service coordinator could possibly be able to utilize the combination between building information modeling and sensor data, the principal of the case school said that it could possibly be helpful to see the whole building at a glance.

The principal of the case school brought our focus to a very simple, but serious problem: how is the basic staff able to analyze complex sensor data? She suggested, that reference values or ranges should be published for all the sensors, so that every data user could understand, when a specific sensor has a value too large or too small.

When asked for thoughts about connecting external data, such as weather data, the principal reported, that the air conditioning system did not respond to weather getting warmer in spring. This means, that there could be room for improvement with predictive weather analytics, which are connected to the air conditioning systems.

The principal of the case school shared our views on utilizing the combination of building information modelling and sensor data in emergency situations. She came up with an example that during a fire emergency, the rooms where the temperature has risen and the rooms where there are still people could be highlighted in order to make the firefighters’ job easier.

4.4.4 The second interview

The second interview was conducted with the energy supervisor of the city. In his own words, he is “the energy expert of real estate in the city”. He is interested in data visualization and data in general. As a part of his job, he developed a real time monitoring system with Power BI software, with which real time sensor data is retrieved from the properties.

In his job, saving potentials can be found from the data. These potentials can be added to the investing list. These potentials can be something like installing solar panels on the roof of a school.

The energy supervisor of the city introduced us to the ”Energy wise cities” (literal translation from Finnish Energiaviisaat kaupungit), which is a project by the six largest cities in Finland in order to achieve energy efficiency together. According to him, large investments are not possible to be made alone and the support of other cities makes it possible.

He emphasized the real-timeness of the data visualizations. He argued, that current data is the most interesting and compelling data of all and that monitoring could be made more efficient with real time data.

According to him, the power of air conditioning units could be lowered, if we could have information about whether the calendar is empty, which would lower the energy consumption. If the temperature in the room rises, the air conditioning could be increased. However, in the optimal situation the system would be predictive instead of reactive, he added.

He said that the consumption per user is a interesting metric, of which Finland’s energy authority is interested in. Before, the consumption per user was not possible to measure, but lately it has become possible, although underutilized, he added. According to him, measuring energy consumption per user is especially useful with low-utilization estates. It is easy to reason, that an empty building should consume as little power as possible.

The energy supervisor of the city noticed a possible security problem: the current amount of people in a school should not be known by many people, let alone everyone. He suggested, that only the principal of the school should be able to see the real-time data of the people counts. He added, that the possibilities of misuse in a security breach are still unknown, so the probability of such event happening should be minimized proactively.

The energy supervisor of the city put emphasis on the data visualization. According to him, maintenance workers are interested in the visualizations and with visualization, the history is interesting too. A good visualization wakes the interest in people, which helps the information spread. *“Kilowatthour does not mean anything in a hour, but during a larger time frame, it becomes meaningful.”* the energy supervisor argued.

One interesting topic during the interview was that data visualization about the energy consumption in the case building was being acquired to his organization during the sprint 2020. However, there hasn’t been any extra information about the current situation of the data visualizations.

The energy supervisor of the city had some insights on the temperature inside the case building: he confirmed some of our hypotheses: the outdoor weather affects indoor temperature and that the warmer air rises up inside a building, which could result in a temperature difference between different floors. Also, he stated that the case building can occasionally consume less energy during the day than at night, because the solar panels on the roof lower the need for external energy during sunny days.

One additional interesting fact, that he told, was that with new buildings, the air conditioning is often kept at maximum power in order to minimize indoor air problems and to dry out the structures. This is often for one year.

4.4.5 The third interview

In the third interview, the service coordinator of the case building was interviewed. He has no permissions to access the sensor data regarding the case building. He described his job description as a modern janitor. When talking about the case building and its data, the service coordinator of the case building stated: “If something has gotten broken over the weekend, someone has to report the incident to me. Only after that I can get to repair. This could be automated. It would be both cheaper and faster.”

When asked about the implementation of such system, he replied that it would be useful to see alarming sensors in the 3d model and watch it with a tablet. He continued: If I had left an door unlocked, or if a door was left open, it would be nice to have an notification and a view in the data model.” According to him, in a case of appliance breakage, the building users should be able to select the room and the appliance from the 3d model. He adds, that the users don’t necessarily remember the number of their room when reporting the breakage, which makes the fixing harder. Also, he adds, it would be nice to able to insert my own comments and then moving the ticket forward.

The service coordinator told, that he does not have any control over the building automation system. He did not see a reason why he would need access to such system, which is a valid question to ask. The discussion moved forward to whether there is any need for him to access it. He concluded, that probably monitoring data could be useful, but full access to the system is not needed.

The service coordinator of the case building has noticed some drawbacks with the current user counter sensor: if two people walk in at the same time, they are counted as one; if there are multiple outdoors in the room, the sensor monitors only one, which results in not every user recorded.

According to him, the fire department could possibly be interested in the people count in the rooms. If fire alarm went off, the exit routes could be shown in the model. However, it is not sure whether such system would be even look at in the case of an emergency.

The service coordinator of the case building also made an interesting argument that the data worker should have some kind of compensation for their additional data work. It has to be remembered, that optimizing some aspect of a whole does not optimize everyone’s processes, but can instead complicate them.

5. RESULTS

5.1 Different themes based on the empirical results

Various views on the possible use-cases regarding sensor data and building information model were found during the interviews. In this chapter, the results from the interviews are gathered and represented. Three main themes were the most dominant in the interviews: efficacy, safety and wellbeing. The occurances of the main themes between interviews are represented in Table 9.

Table 9: Occurance of different discussion themes between interviews

Subject	The principal	The energy supervisor	The service coordinator
Efficacy	True	True	True
Safety	True	False	True
Wellbeing	True	False	False

Occupancy of the building affects largely on the energy performance of a building (Sangogboye et al. 2018). The energy performance is often measured as kWh per user. This means, that the more users there are in the building, given that the energy consumption is constant, the better energy performance is. Also, people radiate heat, which can make the heating more efficient.

In two out of three interviews, the interestingness of the data was mentioned. The principal of the case school and the energy supervisor agreed, that the data about people's own environments is interesting to people. The energy supervisor also emphasized, that real-time data is the most interesting data. If the data is proved to be interesting, the data could be used to create shared understanding on the subject for several people. This could bring people together in new ways, like the pupils of the school studying sensor data together with the principal and a kitchen lady.

Desbite the low overlapping between the interviews, some interesting themes can be observed from the interviews: power optimizing, information sharing / data accesibility monitoring and doubt about staff competence and interests.

The power optimizing with utilizing sensor data and building information models has been greatly under discussion within the 4Apis project and also, the principal of the case school and the energy supervisor of the city had ideas on how to optimize the power usage in the building. The energy supervisor also had an energy perspective to the *Airbnb of the office spaces* idea that the 4Apis had; energy consumption per user is a good indicator for the power efficincy of a facility and it could be lowered by increasing the occupancy rate. The principal of the case school emphasized, that user well-being should be a priority over power savings.

The principal of the case school and the energy supervisor of the city both agreed, that the information should be easy to access and easy to understand by the data consumer. Both of them thought, that the information should be as accessible as possible, without compromising security or privacy. Currently, data security is taken seriously in the case city and the data is not being distributed without proper permissions.

Diversity in the interviewees was desirable in order to obtain a wider spectrum of ideas and viewpoints. This could also explain, why there is not much overlap between the interviews' discussion subjects. This also makes it hard to analyze the results, because the interviews don't validate each other.

5.2 Result utilization in practice

The school spaces are not occupied outside of office hours. This could be optimized by building an information system, which could act as a central hub to bring both the facilities and the groups or individuals that need facilities in the evening time. During 4Apis project's meetings, such groups were identified via expert brainstorming. 4Apis project chose the following groups: people living nearby, associations and community colleges. One use case that came up during brainstorming was that for example in Finland, associations are mandated to have a yearly meeting. Usually, the associations don't have their own facilities and the board members can have their job duties during daytime. Instead of having to rent a meeting room from a hotel or having to meet in a board member's home, they could reserve a classroom from the facility renting service of the city. Second use case would be that different kinds of hobby clubs could be held in the empty school spaces during evenings. During a 4Apis expert brainstorming, it was brought up that this is already being done, but it could be made more systematic.

Since the temperatures in the classrooms rose during the day, it could be possible to optimize the heating and air ventilation systems. Also, lower carbon dioxide levels have been seen to lead to better information usage and strategic plan-making. (Allen Joseph G. et al. 2016) To cycle the carbon dioxide, more power is needed in order to circulate the air with air ventilation system. The middle ground between power usage, temperature and carbon dioxide levels should be researched.

Wifi sensors and cameras could be utilized to enhance the occupancy monitoring system (Sangogboye et al. 2018). This is especially true in the case building, since there are doubts about the reliability of the occupancy sensors as described by the service coordinator of the building and acquired sensor data.

6. CONCLUSIONS

In this chapter, the conclusions of this thesis are discussed. This includes the conclusions from the results of the study, the critique regarding the study and the suggestions for future research. The aim of this thesis was to collect more knowledge about the combination of sensor data and building information models from the perspective of a multi-purpose environment. The research questions that were determined to be answered with this thesis were: *How can sensor data and building information model be utilized a multi-purpose environment?* and *How can sensor data and building information models be combined in practice?*

6.1 Practical use-possibilities for sensor data and building information models

This study managed to bring multiple views to a complex and technical subject. The three themes found from the interviews, efficacy, safety and sustainability contain many use-cases within them. In practise, using sensor data together with building information model could be used to save money on the energy costs, which could benefit the whole organization that operates the building and since the case building is a public building, savings for the building means savings for the whole city.

The combination of sensor data and building information model could be used in order to spread information in a interesting form. This would include sharing information from the building to other stakeholders, such as the city council and to create shared understanding of the surroundings for the primary users of the building. In the case building, this could include creating small study units which could incorporate sensor data in a creative way.

The emergy units, especially the fire department, could use the combination of sensor data and building information models to gather important information about the situation in different parts of the building during a fire. Possibly, occupancy sensors could be utilized in the unfortunate case of external threat; the emergy worker could have a clearer picture of how many people are in the building and where they are located.

6.2 Evaluation of the research

The objectives of this thesis were to provide a way to combine building information modeling with sensor data and acquire interesting view points on the use-cases of that the integration could bring.

Research questions were answered through both literature and empirical research. The main research question concerns the utilization of sensor data and building information models in a multi-purpose environment. Multiple viewpoints were acquired from the interviews and the literature. Apart from answering the research questions, this study had an objective of creating fascinating and useful knowledge about the combination of sensor data and building information models for the 4Apis project.

However, adequacy of the sample size of the interview research is debatable. It could be argued, that with such a small sample size, no conclusions can be drawn. Therefore the results of this study could be considered controversial.

6.3 Reliability and validity of the research

In this chapter, the reliability and the validity of the research are discussed. Reliability concerns how well the research can be repeated and how well it yields the same results (Carmines and Zeller 1979). Validity reflects how well the research represents the phenomenom that it is supposed to represent. (Carmines and Zeller 1979)

From the data it can be seen that the amounts of people in the empirical case rooms decreases steadily between 16 and 22. It could indicate, that the monitoring system starts to manipulate the data, if the amount of people in the room does not change in a while. This makes the validity of the sensor questionable. This, however, does not affect the other parts of the study.

The sample size of the interviews should have been much larger in order to retrieve reliable results. Another option could have been to limit the scope of the study to only include one kind of interviewees, for example teachers, to the interview study.

The viewpoints acquired from the interviews complemented the ones found from literature. This makes the result more reliable, since it is more unlikely that the same result would have been gotten from two completely different sources.

One good question to think about when evaluating a case study is that is the result generalizable to other cases. When concerning this research, the generalization aspect is very debatable. On the other hand, the utilization use-cases, that were found during the interviews, are general, the small sample size makes it scientifically less probable, that the interview results would reflect all the other buildings in existance.

6.4 Suggestions for future research

As mentioned, the sample size of the interviews was only three. It is suggested, that more interviews on the subject are held in order to gather more valuable qualitative information.

In Sangogboye et al. (2018), Wi-Fi sensors, cameras were used in the occupancy detection. Adding these kind of monitoring tools to the case building could prove to be beneficial when trying to detect how occupied the building is. These advanced occupancy detectors should be considered for future research.

The 4Apis project had the idea of having a “Airbnb of Office Spaces”. One of the identified customers for it were associations. Associations should be interviewed in order to gather information about whether there is any need for such a solution. This could, however be more of a market analysis instead of a scientifical research.

More advanced machine learning algorithms should be used in order to find interesting things from the sensor data. There is an ongoing study conducted by University of Helsinki, which utilizes machine learning algorithms in order to derive insights from the 4Apis case buildings’ sensor data.

REFERENCES

Aleksandrov, M., Diakité, A., Yan, J., Li, W., and Zlatanova, S. (2019). “SYSTEMS ARCHITECTURE FOR MANAGEMENT OF BIM, 3D GIS AND SENSORS DATA.” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Copernicus Publications, Vol.4, 3–10.

Allen Joseph G., MacNaughton Piers, Satish Usha, Santanam Suresh, Vallarino Jose, and Spengler John D. (2016). “Associations of Cognitive Function Scores with Carbon Dioxide, Ventilation, and Volatile Organic Compound Exposures in Office Workers: A Controlled Exposure Study of Green and Conventional Office Environments.” *Environmental Health Perspectives*, Environmental Health Perspectives, Vol.124(6), 805–812.

Alves, M., Carreira, P., and Costa, A. A. (2017). “BIMSL: A generic approach to the integration of building information models with real-time sensor data.” *Automation in Construction*, Vol.84, 304–314.

“BACnet Overview.” (n.d.). <http://www.bacnet.org/Overview/index.html> (Sep. 30, 2020).

Bang, M., Engelsgaard, S. S., Alexandersen, E. K., Riber Skydt, M., Shaker, H. R., and Jradi, M. (2019). “Novel real-time model-based fault detection method for automatic identification of abnormal energy performance in building ventilation units.” *Energy and Buildings*, Vol.183, 238–251.

Browne, R. H. (2010). “The t-Test p Value and Its Relationship to the Effect Size and P(X>Y).” *The American Statistician*, Taylor & Francis, Vol.64(1), 30–33.

Buckman, A. H., Mayfield, M., and B.M. Beck, S. (2014). “What is a Smart Building?” *Smart and Sustainable Built Environment*, Emerald Group Publishing Limited, Vol.3(2), 92–109.

Carmines, E. G., and Zeller, R. A. (1979). *Reliability and Validity Assessment*. SAGE Publications.

Chiang, C.-T., Chu, C.-P., and Chou, C.-C. (2015). “BIM-enabled Power Consumption Data Management Platform for Rendering and Analysis of Energy Usage Patterns.” *Procedia Engineering*, Defining the future of sustainability and resilience in design, engineering and construction, Vol.118, 554–562.

Chitu, C., Stamatescu, G., Stamatescu, I., and Sgârciu, V. (2019). “Assessment of Occupancy Estimators for Smart Buildings.” *2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, 228–233.

City of Tampere. (2013). “First 3 years - ECO2 - Eco-efficient Tampere 2020.” 60.

Cummings, S. R., Browner, W. S., and Hulley, S. B. (2013). *Designing Clinical Research*. LWW, Philadelphia.

Djenouri, D., Laidi, R., Djenouri, Y., and Balasingham, I. (2019). “Machine Learning for Smart Building Applications: Review and Taxonomy.” *ACM Computing Surveys (CSUR)*, ACM, Vol.52(2), 1–36.

Epstein, E. (2012). *Implementing Successful Building Information Modeling*. Artech House, Boston.

Fandino, W. (2019). “Formulating a good research question: Pearls and pitfalls.” *Indian Journal of Anesthesia; Belagaum*, Medknow Publications & Media Pvt. Ltd., Belagaum, Belagaum, Vol.63(8), 611–616.

Han, L., Potter, S., Beckett, G., Pringle, G., Welch, S., Koo, S.-H., Wickler, G., Usmani, A., Torero, J. L., and Tate, A. (2010). “FireGrid: An e-infrastructure for next-generation emergency response support.” Elsevier Journal of Parallel and Distributed Computing.

Hartwig, Frederick. (1979). *Exploratory data analysis*. Quantitative applications in the social sciences ; 16, Sage, Beverly Hills.

Howell, S. K., Rezgui, Y., and Yuce, B. (2014). “Knowledge-Based Holistic Energy Management of Public Buildings.” *Computing in Civil and Building Engineering (2014)*, American Society of Civil Engineers, Orlando, Florida, United States, 1667–1674.

International Organization for Standardation. (2018). “Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM). Information management using building information modelling. Part 1: Concepts and principles (ISO 19650-1:2018).” *ISO*, <https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/06/80/68078.html> (Jun. 12, 2020).

Israel, M., and Hay, I. (2020). *Research Ethics for Social Scientists*. London.

Kastner, W., Neugschwandtner, G., Soucek, S., and Newman, H. M. (2005). “Communication systems for building automation and control.” *Proceedings of the IEEE*, Vol.93(6), 1178–1203.

Ketokivi, M., and Mantere, S. (2010). “Two Strategies for Inductive Reasoning in Organizational Research.” *Academy of Management Review*, Academy of Management, Vol.35(2), 315–333.

Kunz, J., Fischer, M., Haymaker, J., and Levitt, R. (2002). “Integrated and automated project processes in civil engineering: Experiences of the Center for Integrated Facility Engineering at Stanford University.” *Computing in Civil Engineering Proceedings, ASCE, Reston, VA*, Citeseer, 96–105.

Law, J. L. (2016). “brainstorming.” *A Dictionary of Business and Management*, J. Law, ed., Oxford University Press.

Lee, A., Wu, S., Marshall-Ponting, A., Aouad, G., Cooper, R., Tah, J., Abbott, C., and Barret, P. (2005). *nD Modelling Roadmap: A Vision of the Future of Construction.*

Leppäkoski, H. (2015). “Novel Methods for Personal Indoor Positioning.” Tampere University of Technology.

Lim, C., Kim, K.-H., Kim, M.-J., Heo, J.-Y., Kim, K.-J., and Maglio, P. P. (2018). “From data to value: A nine-factor framework for data-based value creation in information-intensive services.” *International Journal of Information Management*, Vol.39, 121–135.

Petrova, E., Pauwels, P., Svidt, K., and Jensen, R. (2019). “Semantic data mining and linked data for a recommender system in the AEC industry.” 172–181.

Riaz, Z., Parn, E. A., Edwards, D. J., Arslan, M., Shen, C., and Pena-Mora, F. (2017). “BIM and sensor-based data management system for construction safety monitoring.” *Journal of Engineering, Design and Technology*, Emerald Publishing Limited, Vol.15(6), 738–753.

Rio, J., Ferreira, B., and Poças-Martins, J. (2013). “Expansion of IFC model with structural sensors.” *Informes de la Construccion*, Consejo Superior de Investigaciones Científicas, Vol.65(530), 219–228.

Rowley, J. (2007). “The wisdom hierarchy: representations of the DIKW hierarchy.” *Journal of Information Science*, SAGE Publications Ltd, Vol.33(2), 163–180.

Sangogboye, F. C., Arendt, K., Jradi, M., Veje, C., Kjærgaard, M. B., and Jørgensen, B. N. (2018). “The impact of occupancy resolution on the accuracy of building energy performance simulation.” *Proceedings of the 5th Conference on Systems for Built Environments*, ACM, Shenzen China, 103–106.

Saunders, M. N. K., Lewis, P., and Thornhill, A. (2019). *Research Methods for Business Students*. Pearson Education.

Sharma, H., Vaidya, U., and Ganapathysubramanian, B. (2019). “Estimating contaminant distribution from finite sensor data: Perron Frobenious operator and ensemble Kalman Filtering.” *Building and Environment*, Vol.159, 106148.

Sitra. (2020). “Construction.” *Sitra*, <https://www.sitra.fi/en/topics/construction/> (Oct. 28, 2020).

Smaoui, N., Kim, K., Gnawali, O., Lee, Y.-J., and Suh, W. (2018). “Respirable Dust Monitoring in Construction Sites and Visualization in Building Information Modeling Using Real-time Sensor Data.” *Sensors and Materials*, Vol.30(8), 1775.

Smith, I. F. C. (2016). “Studies of Sensor-Data Interpretation for Asset Management of the Built Environment.” *Frontiers in Built Environment*, Frontiers Media S.A., Vol.2.

Snoonian, D. (2003). “Smart buildings.” *IEEE Spectrum*, Vol.40(8), 18–23.

Tang, S., Shelden, D. R., Eastman, C. M., Pishdad-Bozorgi, P., and Gao, X. (2020). “BIM assisted Building Automation System information exchange using BACnet and IFC.” *Automation in Construction*, Vol.110, 103049.

The Institution of Structural Engineers. (2020). “BIM guidance part 1: Introduction to BIM - The Institution of Structural Engineers.” <https://www.istructe.org/resources/guidance/bim-guidance-part-1-introduction-to-bim/> (Jun. 12, 2020).

Vieira, R., Carreira, P., Domingues, P., and Costa, A. A. (2020). “Supporting building automation systems in BIM/IFC: reviewing the existing information gap.” *Engineering, Construction and Architectural Management*, Emerald Publishing Limited, Vol.27(6), 1357–1375.

Yang, L., Yan, H., and Lam, J. C. (2014). “Thermal comfort and building energy consumption implications – A review.” *Applied Energy*, Vol.115, 164–173.

Yaniv, Z., Link to external site, this link will open in a new window, Lowekamp, B. C., Johnson, H. J., and Beare, R. (2018). “SimpleITK Image-Analysis Notebooks: a Collaborative Environment for Education and Reproducible Research.” *Journal of Digital Imaging; New York*, Springer Nature B.V., New York, Netherlands, New York, Vol.31(3), 290–303.

APPENDIX A: ZEROPAD FUNCTION

```
def zeropad(x):
    """Adds leading zeros to non-zero padded
    datetime values.
    Parameters:
    x -- datetime-like string in 0.0.2019 00:00:00 format
    """
    s_x = x.split('.')

    # Add the zero to the day
    if len(s_x[0]) == 1:
        s_x[0] = "0" + s_x[0]

    # Add the zero to the month
    if len(s_x[1]) == 1:
        s_x[1] = "0" + s_x[1]

    # Join the splitted array into a string and return
    return '.'.join(s_x)
```

APPENDIX B: THE INTERVIEW FRAME

4APIs
THE VALUE FOR THE USER, BIM + SENSOR DATA, 4/2020

Interview frame, 60 min.

INTRODUCTION (10 min.)

- The principles of the interview
  - the usage of the materials, GDPR (n.b. notes, recording, to support research and usage) – form
- The characteristics and goals of the interview
- Project background
- Introduce oneself

ROLE, JOB AND ORGANIZATION (15 min.)

Let's talk in more detail about the interviewee's…
- organization
- role
- job description
- A typical day, everyday processe
  - When have you been successful (in relation to your own / role's objectives)? And more broadly in terms of your organization (not just your own role)?
  - What are the challenges / obstacles to success?
- To what extent do you deal with a) data b) data models in your work? What kind?
  - What kind of experiences?
- How much do you have to do with the case building? In what matters?
  - What kind of experiences?

4 APIs, building information models & data (25 min.)

Discuss the project, opportunities and describe them in more detail if necessary…

- Are you familiar with the 4APIs project? If so, what did you know about it before this interview?
- What are your thoughts on the project?
- What are your thoughts on building-bound data in general, its possibilities?
  - Positives – To whom is it important and why? → What does it require to be realized?
  - Negatives – To whom is it negative and why? → Could the risk be somehow anticipated, minimized, eliminated?
- What are your thoughts on the data tied to the case building, about the posssibilities of it?

   - Positives – To whom is it important and why? → What does it require to be realized?
   - Negatives – To whom is it negative and why? → Could the risk be somehow anticipated, minimized, eliminated?

Let’s go through a few drawings, hypotheses, and illustrations (separate material) of potential value for different users. Each is discussed…

- What are the first thoughts about this idea?
   - Positives – To whom is it important and why? → What does it require to be realized?
   - Negatives – To whom is it negative and why? → Could the risk be somehow anticipated, minimized, eliminated?
   - What should be done next with this idea? What direction is should be taken? What to change and why?

If not already brought up, these should be addressed:

- Indoor air solutions: Do you have information about optimizing air quality in other locations?
- The thematics of ”saving”: What exactly is the opportunity to save here? How are they balanced?
   - Financial? How significant savings?
   - Healthy? How significant savings?
   - Ecological? How significant savings?
   - other?
- How could combining BIM and sensor data contribute to the goals of the Carbon Neutral City project?
   - The possibilities discussed here?
   - Other?

Finally (5 min.)

- What is the main thing that is on your mind about this interview?
- What feels the most promising? What are the challenges to consider? What next?
- Any other greetings to the project group?